0782

PB 204 193

COMPUTER MAPPING
as an aid in
AIR POLLUTION STUDIES

VOLUME TWO: INDIVIDUAL REPORTS

REPORT G
SOUTHERN NEW ENGLAND STUDY

by
William Warntz
John Goodrich
and
Steven Fosburg

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

A

56

BIBLIOGRAPHIC DATA SHEET	1. Report No. APTD-0782	2.	3. Recipient's Accession No.
4. Title and Subtitle Computer Mapping as an aid in Air Pollution Studies Volume Two: Individual Reports - Report G - Southern New England Study			5. Report Date April 1970 6.
7. Author(s) William Warntz, John Goodrich and Steven Fosburg			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Graduate School of Design Harvard University Cambridge, Massachusetts 02138			10. Project/Task/Work Unit No. 11. Contract/Grant No. 68A-2405D
12. Sponsoring Organization Name and Address National Air Pollution Control Administration U. S. Public Health Service U. S. Department of Health, Education, and Welfare Washington, D. C. 20201			13. Type of Report & Period Covered 14.

15. Supplementary Notes DISCLAIMER - This report was furnished to the Office of Air Programs by Graduate School of Design, Harvard University, Cambridge, Massachusetts 02138 in fulfillment of Grant 68A-2405D

16. Abstracts

A report is presented of a study dealing with computer mapping and the air pollution problems of Southern New England. The study presents an introduction discussing macrogeology and field quantity theory, and income potential as an index of air quality. The second section dealing with application of the study to New England data, includes discussions of air pollution surfaces, socio-economic surfaces and air quality control regions. An appendix describing a program to determine the critical features of a surface is included.

17. Key Words and Document Analysis. 17a. Descriptors

Air pollution
Computer programs
Mapping
Geography
Income
Data analysis
Socioeconomic status

17b. Identifiers/Open-Ended Terms

New England
Air Quality Control Regions

17c. COSATI Field/Group 13B

18. Availability Statement Unlimited	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 156
A	20. Security Class (This Page) UNCLASSIFIED	22. Price 3.00-0.95

USCOMM-DC 4[illegible]-P71

LABORATORY FOR COMPUTER GRAPHICS
AND SPATIAL ANALYSIS

Graduate School of Design
Harvard University
Cambridge, Massachusetts

This report was funded by:

Demonstration Grant 68A-2405D
National Air Pollution Control Administration
U. S. Public Health Service
U. S. Department of Health, Education, and Welfare

John C. Goodrich, Editor
Gail Howrigan, Associate Editor

Rosalie Alexander
Carol Chase
Candace Howes
Whitmore John
Linda Peacock

April 1970

COMPUTER MAPPING

as an aid in

AIR POLLUTION STUDIES

Personnel

Principal Investigators

Howard T. Fisher	Project Director (Associate Director of The Laboratory for Computer Graphics and Spatial Analysis)
John C. Goodrich	Associate Project Director
William Warntz	(Director of The Laboratory for Computer Graphics and Spatial Analysis)
Allan H. Schmidt	(Associate Director of The Laboratory for Computer Graphics and Spatial Analysis)

Case Study Investigators

Norma Bogen	St. Louis Project
Paul Jack Dangermond	California and Puget Sound Studies
Steven Fosburg	Statistical Investigation
Jean Maurice Granger	Montreal Project
Donald Shepard	St. Louis Project/Program Investigation
Elia Werczberger	Kansas City Project

Programming Support and Development

Chang-I Hua
Lannon Leiman
Kathleen Reine

Research Assistants

Whitmore John
Johnny Kirkland
Helen Mansfield
Adrian Thomas

Ellen Bussard
Robert Cartwright
Katherine Mickey
Warnie Webster

Administrative and Secretarial Support

Malcolm Fitzpatrick
Gail Howrigan
Linda Peacock
Laurence Yont

Regina McDonald
Donna L. Moorhead
Claudia S. Rosen
Judith Strong

PREFACE

From July 1, 1968 through December 31, 1969, the Laboratory for Computer Graphics and Spatial Analysis conducted a demonstration of several conceptual, methodological, and technical approaches to problems in the field of air pollution, employing computer-produced maps and allied displays. This work was done under a grant (No. 68A-2405D) from the Division of Control Agency Development, National Air Pollution Control Administration (NAPCA), Public Health Service, U. S. Department of Health, Education, and Welfare.

The results of this study, taken in its entirety, should make an important contribution to the work of Federal, state, and local air pollution agencies, especially in their efforts to determine the best control strategies for use in air quality control regions. The computer is an ideal tool for relating and combining sets of data, and for considering alternatives. For instance, graphic displays can contribute significantly to the evaluation of both short-term and long-term control strategies.

In demonstrating "the use of computer graphics as a means for more clearly portraying data related to the sources, distribution, and community effects of air pollution," the primary emphasis was on case studies for specific areas. The secondary emphasis, to support and complement the case studies, included:

(a) "the development of instructional manuals covering the procedures required for the preparation and interpretation of computer maps of various kinds, as related to air pollution,"

(b) "the use of certain newly developed techniques for describing, analyzing, and portraying quantitative surfaces and movements relative to such surfaces, as related to air poliution," and

(c) "an appraisal of alternative equipment, techniques, and computer programs presently or potentially available for the production of computer graphic displays pertinent to air pollution."

The case study regions were selected in consultation with NAPCA. Considerations included:

(a) sufficient air quality, emission, and related socio-economic and meteorological data--spatially distributed over the region--to enhance the likelihood of adequately demonstrating the mapping techniques;

(b) representative problems in data handling and mapping so that the techniques demonstrated would have applicability to as many other areas as possible;

(c) special problems in data handling and mapping so that as many techniques as possible could be adapted, developed, and tested;

(d) areas of which NAPCA and the individual investigators had special knowledge and could therefore utilize relevant research to enhance the study;

(e) areas where an immediate interest in mapping had been expressed for on-going Federal, state, and local air pollution activities.

Other research was undertaken to support and complement the case studies, and, in particular, to provide:

(a) the widest possible distribution of the mapping programs and accompanying explanation of the techniques through

i) copies of the Laboratory's SYMAP Computer mapping program,
ii) an introductory correspondence course in the SYMAP program,
iii) general program description and supplementary explanation for the existing user-manuals;

(b) a general review of present and potential hardware, output devices, and programs applicable to computer graphic displays of air pollution and related data;

(c) means for predicting, projecting, and correlating air pollution and related data, utilizing the latest

knowledge in planning, geography, and spatial analysis in general;

(d) dissemination of information on approaches and techniques to Federal, state, and local air pollution agencies, especially for use in the air quality control regions.

The final report is divided into two parts: Volume One: Summary, and Volume Two: Individual Reports--the detailed reports on the case study areas and the technical reports related to other research undertaken. The pertinent graphic displays are included with the appropriate individual report and a complete set is provided with the Summary Volume.

Introductory sections describing the air pollution and related data used are provided in the major case study reports for St. Louis and Kansas City; these should be referred to by anyone unfamiliar with the types of data and problems existing in this field. Detailed discussions of special data handling and mapping techniques are presented either as appendices to case study reports, or in separate technical reports.

The Summary Volume includes a discussion of all the areas of research undertaken and contains the complete table of contents for the reports of Volume Two. This should provide a useful reference of the special problems and technical discussions of one report which may be of interest to the reader of another detailed report.

The Southern New England study is presented here as Report G of Volume Two. It has a section on air pollution surfaces and socio-economic surfaces for New England, in addition to an introduction discussing macrogeography and field quantity theory, and income potential as an index of air quality. An appendix describing a program to determine the critical features of a surface is included, in addition to the appendix of computer graphics produced.

The data used in this study was provided by Sales Management, Inc., New York, the Commonwealth of Massachusetts, Department of Public Health, and the State of Connecticut, Department of Health. Estimates of air quality were made in conjunction with personnel of the Commonwealth of Massachusetts, Department of Public Health.

The study was under the direction of William Warntz, with the assistance of John Goodrich and Steven Fosburg. The report was prepared by the investigators.

John C. Goodrich, Editor
Cambridge, Massachusetts
April, 1970

REPORT G
SOUTHERN NEW ENGLAND STUDY

Table of Contents

PREFACE

1 INTRODUCTION

1.1. Macrogeography and Field Quantity Theory

1.2. Income Potential as an Index of Air Quality

1.2.1. Northeast Corridor
1.2.2. National Atmospheric Areas

2. APPLICATION TO NEW ENGLAND DATA

2.1. Air Pollution Surfaces

2.1.1. Available Sources of Data
2.1.2. Estimating Procedures

2.2. Socio-Economic Surfaces

2.2.1. Calculating Potentials
2.2.2. Correlation Studies
2.2.3. Critical Features of Surfaces

2.3. Air Quality Control Regions

REFERENCES

APPENDIX 1 COMPUTER GRAPHICS
APPENDIX 2 CRITICAL FEATURES PROGRAM

A. Main Features of the Program
B. Program Requirements

List of Illustrations

1. INTRODUCTION

Example:

1-1 Modeled Carbon Dioxide in Northeast Corridor

1-2 Income Potential Surface for Northeast Corridor

2. APPLICATION TO NEW ENGLAND DATA

Example:

2-1 Air Pollution Data Sources

APPENDIX 1 COMPUTER GRAPHICS

Figure:

G-1 Concentration of Suspended Particulates Based on 206 Selected Urban and Rural Recordings
- a. Equal Number of Data Points in Each Level
- b. Equal Interval Contour Levels

G-2 United States Income Potential Based on 206 Urban and Rural County Values
- a. Equal number of Data Points in Each Level
- b. Equal Interval Contour Levels

G-3 Concentration of Suspended Particulates
- a. Base Map of Monitoring Stations and Background Data Points
- b. Based on Actual and Estimated Values

G-4 Income Potential Calculated at City & County Data Points
- a. Surface Based on Counties Only
- b. Surface Based on Cities and Counties

G-5 Concentration of Suspended Particulates
- a. Based on Measured 1966 Values Only
- b. Based on Actual and Estimated Values

G-6 Income Potential Surface
- a. Calculated at City and County Data Points
- b. Oblique View of a.
- c. Calculated on an 8-mile Grid
- d. Oblique View of b.

G-7 Comparison of Income Surfaces--Oblique Views
- a. Income Density
- b. Income Potential at Data Points
- c. Income Potential on an 8-mile Grid

G-8 Comparison of Population Surfaces
- a. Population Density
- b. Population Potential at Data Points
- c. Population Potential on an 8-mile Grid

G-9 Comparison of Air Pollution and Income Potential Surfaces: Oblique Views
- a. Concentration of Suspended Particulates: Actual and Estimated Values Used
- b. Income Potential--Calculated on an 8-mile Grid

G-10 Critical Features of Income Potential Surface
- a. Income Potential Surface
- b. Hills and Course Lines
- c. Dales and Ridge Lines

1. INTRODUCTION

A number of important studies[1] have determined that good correlations can be found between socio-economic variables and spatial configurations. For example, many quantities vary according to the ratio of population divided by distance: P / r

If for any location or point, in an area or region, the population divided by the distance to that point--for all other locations having a population--is summed, an "index of influence of population at a distance" is found for that location. Repeating this procedure so as to calculate the index of influence for all other possible locations, a surface is generated for the area; this surface is termed a "population potential surface". Further, if each population is weighted by its corresponding income, then an "income potential surface" is calculated. The income potential surface correlates well with numerous quantities and often gives better results than the population potential surface.

Based on the results of these previous studies, it was hypothesized that the surfaces of potential could be used as an indicator of overall air quality for a region. Potentials are an aggregate measure of the concentrations of people and their wealth; air pollution, for the most part, is also a good indication of the concentrations of people and their activities, or 'wealth'.

1.1. MACROGEOGRAPHY AND FIELD QUANTITY THEORY

A potential of population surface may be regarded as revealing a "socio-economic terrain." As a spatially continuous surface, it shows the structuring for exchanges relating to social and economic phenomena. Potential, itself, is a measure

of influence at a distance or of "macroscopic aggregate accessibility" to a population. When we consider the entire population distribution of a country, not as groups about points, but, rather as a density distribution in the plane, then the total potential of population at any given point c may be computed as:

$$V_c = \frac{1}{r} GdA.$$

GdA signifies the density of population, G, over any infinitesimal element of area, dA, and r the distance of each said element to the given point c. The integration is extended to all areas of the plane where G is not zero.

If we substitute D (income density) for G, and perform the same calculations we obtain the income potential at a point: U_c. Integration for all areas where D is not zero determines the income potential surface. Such surfaces can be thought of as showing the "aggregate spatial accessibility or influence" of population or income for the country.

Repeated studies have demonstrated very remarkable correlations for the spatial patterns of social and economic phenomena with this potential of population surface for the United States. Constancy in certain phenomena is exhibited along the contours of equipotential and systematic variation in these phenomena across the potential field. For example, the following were found:

-- Rural population density varies as the square of the potential;

-- Rural nonfarm population density as the cube;

-- Rural nonfarm rent varies as the first power;

-- Land values vary as the square;

--Railroad track density varies as the first power;

--Rural free delivery route density varies as the three-halves power;

--Density of wage earners in manufacturing varies as the fifth power; and,

--Death rates vary as the square root of the potential.

Other similar relationships have been established with respect to telephone network density, rail-line density, road density, market area, administrative areas, farm sizes, alcoholism, mental disorders, the "velocity" of money, bank check clearances, first class mail, patents, telephone calls, and so on. Potential of population can, therefore, be thought of as a measure of "sociological intensity"--the rate at which decisions are made and information is processed.

Potential of population or income is a simple integrative macroscopic index of the influence of people or wealth at a distance. Individual "microscopic" behaviors of a given sort average to yield a "macroscopic" behavior-type of the simplicity to fit a numerical index. This facility of substitution--a behavior-type for behavior--makes potential both a highly simplifying and a strongly unifying concept. The idea of potential carries with it a field quantity theory--a system of related factors (gradient, "energy," population density), each of which projects two basic "dimensions"; population and distance. It may be useful to think of total influence of population in a different way:

(i) Consider that each person's influence is represented by a pile of sand, with the highest point of the pile at the place the person occupies and decreasing heights with increased distance from the person.

(ii) Suppose that there is a similarly constructed sandpile around the place of residence of every individual.

(iii) Now, let all these sandpiles be superimposed; at any point the total height of the sand will be the sum of the heights of all the individual sandpiles.

(iv) Let this height be a measure of total population influence at that point; then, a contour map or a physical model may be made of the entire surface by connecting all of the points.

Even when the idea of potential of population was first presented, it was clear that not all people could be assigned equal "weights" in the formula for influence. Not every person can be regarded as the "standard American", in terms of his ability to attract, interact, and influence at a distance.

The idea was advanced that the "social mass per capita" in an area represented the appropriate "weight". Whatever is artificially produced or is transported for social purposes is social mass. The physical mass of artificial goods required to support and sustain modern civilization is great indeed; in the United States the per capita mass of artificial goods amounts to two or three thousand tons, including such things as houses, railway and highway roadbeds, public buildings, automobiles, bridges, dams, water pumped into aqueducts, and locomotives.

The amount of social mass varies throughout the world, but statistics indicating the per capita mass of artificial goods in different places are not readily available. An acceptable substitute for social mass might be the weighting of people by income to produce income potential.

1.2. INCOME POTENTIAL AS AN INDEX OF AIR QUALITY

The relationship of population and income potential to air pollution depends, primarily, on certain heuristic assumptions which have been borne out by the general similarity of the potential and air pollution surfaces. Statistical analysis of the air pollution data was attempted to see if the two surfaces were as related as was evident by the first approximations. We can start with some function of the potential and then seek the socio-economic interrelatedness of the systems.

We are limited, however, because the air quality data available for the study are fundamentally unsuited for the purposes. They were used simply because there were no other data available. Consequently, our results must be considered as a demonstration of techniques at this point, rather than as a statistical test of the relationship of the two kinds of surfaces.

The air quality data are inadequate for several reasons. First, each air quality sampling point represents mainly the local activity of the air pollution surface within the vicinity of the measuring station. The readings are very much influenced by local emissions of particular stationary sources, by mobile source patterns, and by the particular wind patterns. Moreover, there are too few sampling stations when we look at the country as a whole to sufficiently check possible errors in measurement. A high value found for a small area may not adequately represent that area in general.

It is difficult to derive an air pollution index. The potential surface should reflect all major pollutants and, therefore, an appropriate air pollution index; an average air pollution indicator.

It is very difficult to actually qualify "average", either in a time sense or in a spatial sense. We have attempted to compare potential only for one year and with one pollutant, suspended particulates.

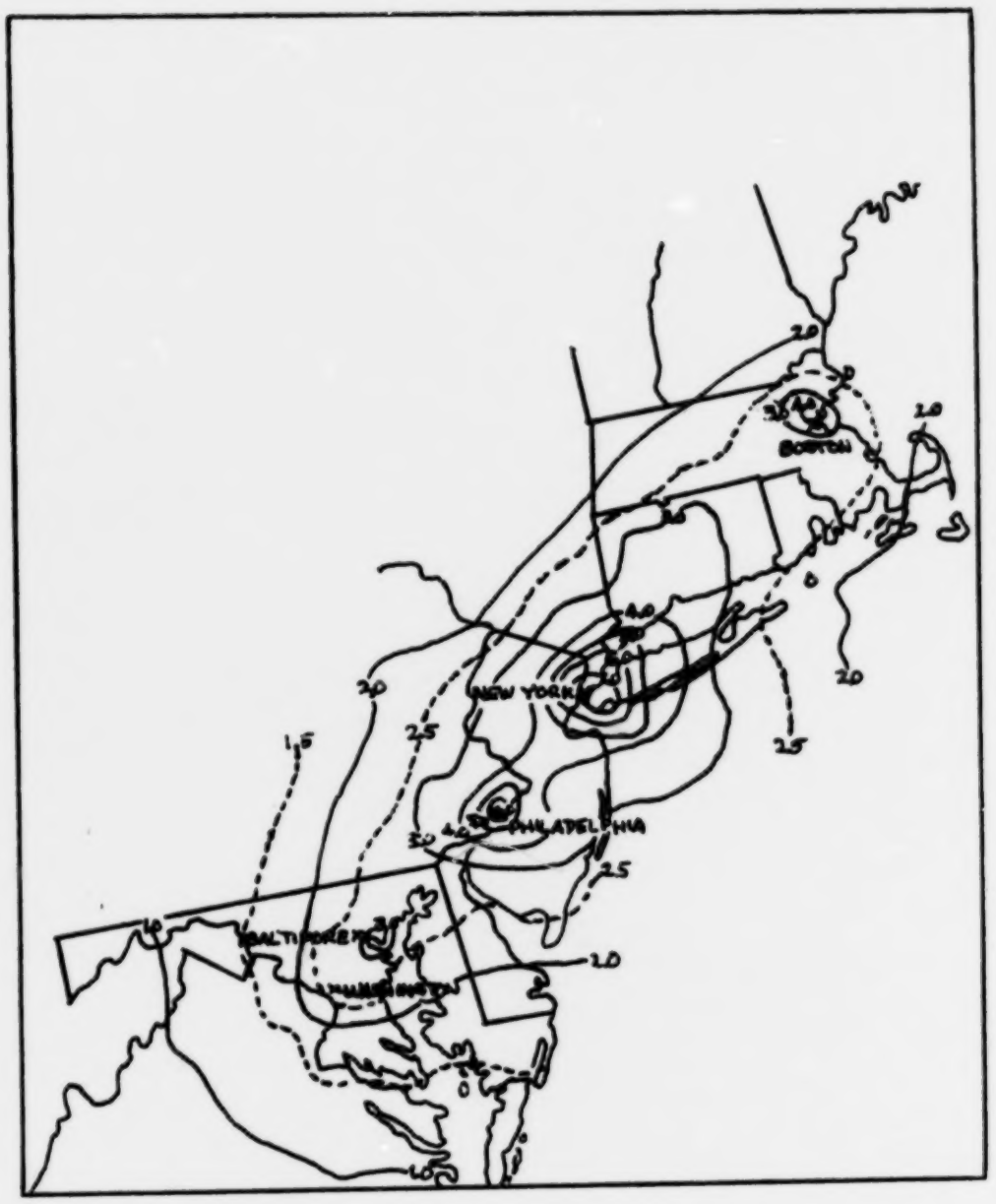
1.2.1. Northeast Corridor

Preliminary investigation of the correlation of air quality with the income potential surface suggests that, when an empirically determined factor of proportionality is used, potential values can serve as a first approximation for predicting air quality. Such an index essentially reflects area source emissions and should be supplemented by other information, such as point emission and meteorological data.

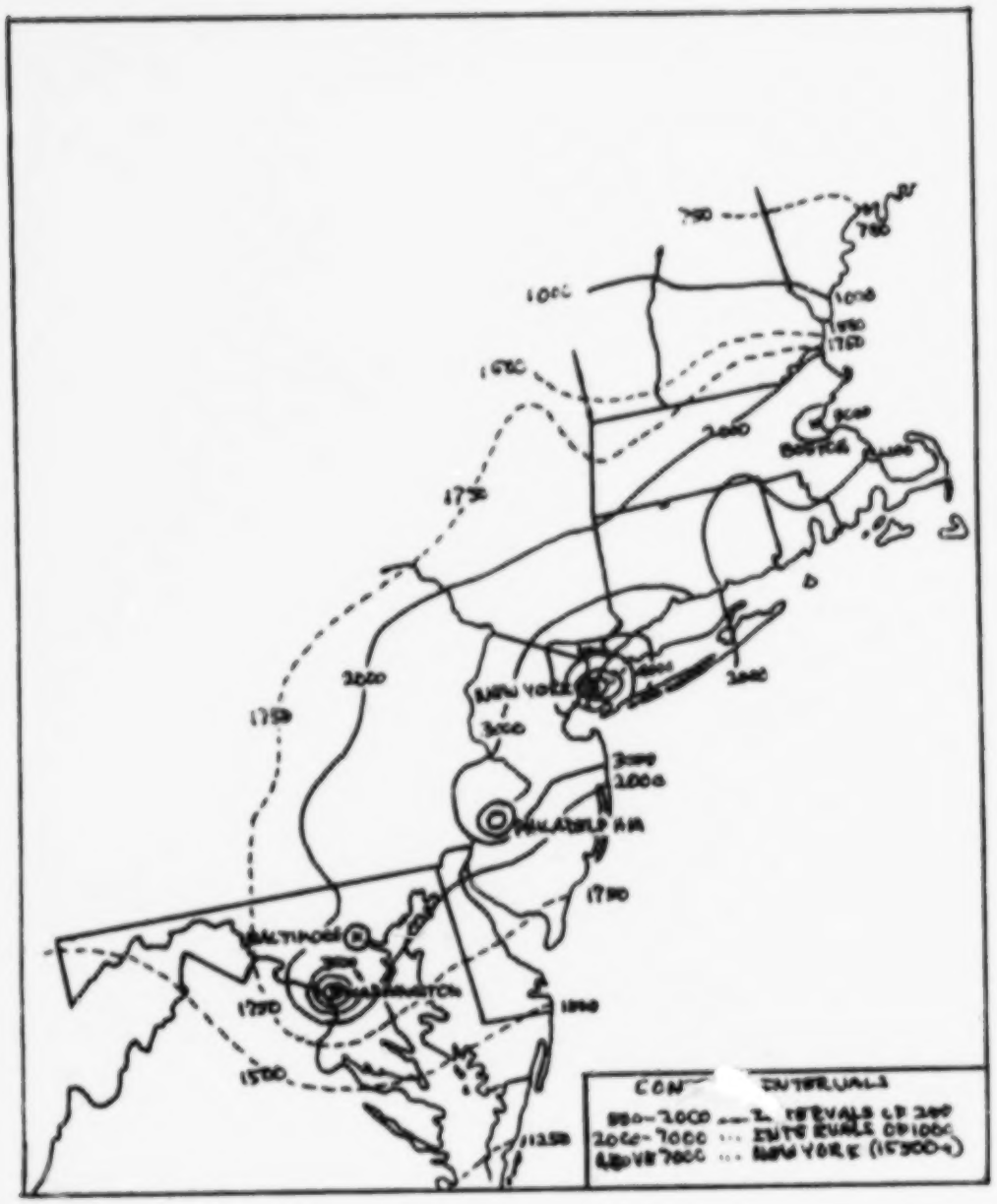
The first correlation of air quality and potential was made for data for the northeast corridor from Boston to Washington. The 1967 United States income potential mapping shown in Example 1-2 was derived from county income data[2]; the 1967 modeled carbon dioxide values (infinite half life) shown in Example 1-1 were also derived from estimated county CO_2 emissions.[3] Graphically, the isopleth patterns are very similar.

A coefficient of rank correlation based on concomitant values for a sample of 60 randomly selected control points on the surfaces was computed as 0.91. A graphing on double logarithmic paper revealed that air quality, as indicated by the Carbon Dioxide Index, varied approximately as the first power of the Income Potential.

This was by far the most encouraging correlation found between the two kinds of surfaces in our studies. However, the data bases used here were also the only reliable ones found:



Example 1-1
Modeled Carbon Dioxide in Northeast Corridor



Example 1-2
Income Potential Surface for Northeast Corridor

(1) County values were used in both cases; generally, air quality is derived from point and area data.

(2) Modeled air quality was used, rather than measured; this is much more apt to give a smoothed surface, similar in nature to the potentials surface. Such a smoothed surface eliminates the local measuring point errors and may give a truer representation of air quality variations on the macroscale.

(3) An area source type pollutant was used. This would be more apt to reveal a smoothed surface then a pollutant such as suspended particulates, which results in large measure from point sources.

(4) A predominant wind along the axis of population reinforces the predominant socio-economic potentials interaction on this axis. A southeast wind, for instance would have given a poor correlation.

1.2.2. National Atmospheric Areas

Sec. 107(a)(1) of the Air Quality Act of 1967 stated that:

"The Secretary shall, as soon as practicable, but not later than one year after the date of enactment of the Air Quality Act of 1967, define for the purposes of this Act, atmospheric areas of the Nation on the basis of those conditions, including, but not limited to, climate, meteorology, and topography, which affect the interchange and diffusion of pollutants in the atmosphere."(4)

On January 9, 1968 Secretary John W. Gardner announced the definition of eight atmospheric areas determined "on the basis of those conditions which affect the interchange and diffusion of pollutants in the atmosphere." These areas were based mainly on annual averages of meteorological parameters, such as: (i) low-level inversion frequency, (ii) maximum depths of vertical mixing, and (iii) the frequency of light winds. The boundaries were broad bands, rather than detailed lines, and reflected the general annual-integrated meteorology, rather than the meteorology in the vicinity of the boundary or at a specific time.

We wished to compare (1) income potential and air pollution data for the country as a whole, and (2) given a reasonably good correlation, use the income potential surface to define the equivalent of national atmospheric areas.

A map of the United States, with data points at the centroid of each of the over 3000 counties, was used for this study(5) Data from the National Air Sampling Network (NASN) for 1967(6) was available for suspended particulates at 206 selected urban and rural points. These points were assigned to the centroid of the appropriate county and mapped. Figure G-1 shows the resultant surface of national air quality for two different contour intervals. The number and spacing of these 206 data points is clearly inadequate to give even a macroscale picture of air quality; in addition, the mixture of urban and rural readings skews the results.

The income potential for the same 206 out of 3000 counties was mapped and is shown in Figure G-2 for the same contour intervals used in Figure G-1. This mapping clearly revealed

the small number and biased spacing of data point locations. With such insufficient data for comparison, no further mapping was done at the national scale.

A subset of these 206 rural data points was selected for statistical (but not graphical) comparison; the data points all represented rural settings where "background values" of air pollution might be expected. Concomitant values for 32 rural control points widely scattered over the United States correlated at the 0.5 level. The air pollution index, as the dependant variable in this case, was the geometric mean of the observations for 1964-65 for suspended particulates (micrograms per cubic meter) and the independant variable was the income potential surface for 1967. Many of the air pollution observations were taken in National Parks and hence may not be representative of the variation among rural areas.

Although no further work could be attempted at the national scale due to lack of data, these preliminary results lend encouragement to the future application of potentials surfaces for macroscale air pollution indices.

2. APPLICATION TO NEW ENGLAND DATA

The principle study area chosen for the examination of air quality and potentials surfaces was southern New England; the area was chosen because it had the best combinations of available air quality and potentials data. Studies of population and income potential had already been undertaken[7] and air quality data for suspended particulates was available from several sources.

The data was not sufficient to undertake meaningful statistical correlations. Consequently, only mapping and visual comparisons were done. Again, the results were encouraging, leading to the conclusion that significant comparisons can be made once good networks of monitoring stations, or reliable modeling techniques, exist. When this is the case the potentials surfaces can be used to quickly and inexpensively estimate future macroscale levels of air quality.

2.1. AIR POLLUTION SURFACES

Suspended pariculates is the only pollutant for which a sufficient amount of reliable data was available for southern New England. Data for 1966 and 1968 were used, together with updates from previous years and estimates of "background values" for areas where no data were known. This conglomeration of data sources was used to determine the air quality surface for comparison with the 1967 income potential surface.

2.1.1. Available Sources of Data

Example 2-1 summarizes the air pollution data sources available for this project.[8] Three main sources of data were used: (i) Federal Continuous Air Monitoring Network (CAMP) and National Air Sampling Network (NASN) measurements for 1966, (ii) state and city readings in Massachusetts and Connecticut for 1966 and 1968, and (iii) estimated and updated values for "background" areas. These sources of data were used for the study area comprising the three southern New England states and for the surrounding areas of New York, New Hampshire, Vermont, and Maine so as to estimate values on the periphery of the study area. The data points at the periphery minimize the erroneous extrapolation of values often found at the edges of the study area with computer mapping.

The most spatially comprehensive and detailed data were available for the Boston area (20 data points) and the state of Connecticut (24 points). This was supplemented with Federal and State data at 9 locations in Massachusetts, 5 in Connecticut, and 2 in Rhode Island. The Rhode Island data is particularly weak: one station was in downtown Providence and the other

Data Point Locations and Sources of Data

	Southern New England				Background	
	Massachusetts	Connecticut	Rhode Island	Sub-T	New York & N. New England	Total
Federal Data				11		40
1966 Camp Urban	4	5	1		26	
1966 Camp Rural			1		3	
State Data				49		50
1968-Connecticut		24				
1966-Boston area	20					
1966-NASN(8 Mo.)	2				1	
1968-Mass. Cities	3					
Sub-Total	29	29	2	60	30	90
Estimated Data						
Updated '54-'62	4				1	
background estimates	23					
Total	56	29	2	87	31	118

Example 2-1
Air Pollution Data Sources

in a state forest. This gives a total of 60 data point locations in the study area for actual 1966 and 1968 values. For mapping, this was supplemented with 30 actual values at data point locations in surrounding states and 28 estimated or updated values, 27 of them in the study area itself.

2.1.2. Estimating Procedures

Two procedures were used to locate the additional 28 points for mapping, and to determine the appropriate values for suspended particulates. For five towns air quality had been measured during earlier years. These data point locations were used and the values 'updated' to 1968: the values were kept the same unless there was some reason to expect a significant increase or decrease. The locations included 4 towns in Massachusetts with readings which varied from 1954 to 1962 and one town in New Hampshire with 1954 readings.

Significant portions of Massachusetts, in particular, had no available data. Maps using the available 95 data points (90 measured plus 5 updated) revealed the erroneous impression given by the data point placement (see Figure G-5). Accordingly, at 23 strategically located points in Massachusetts "background values" were assigned based on the actual readings found for similar areas, such as suburban and rural Boston and Nantucket. The 'threshold' background value for this region appeared to be 40 micrograms per cubic meter for the annual average of daily readings.

It is evident from the base map in Figure G-3 that the data point spacing is inadequate in parts of Connecticut and most of Rhode Island; this affects the reliability of the mapping in

these areas. Unfortunately, it was not possible to make reasonable estimates of background values for these areas, as was done in Massachusetts.

The surfaces developed from these 118 data points appeared to be more meaningful, at least in the number and spacing of data point locations (see Figures G-3 and G-5). To develop more reasonable surfaces of air quality for correlation with income potential and other variables, measurements will have to be taken at many sites and the procedures for estimating and testing background values will have to be refined. In the meantime advection-diffusion modeling can be employed to estimate air quality when a good emission inventory is available; no such inventory has been prepared for most of the study area.

2.2. SOCIO-ECONOMIC SURFACES

Efforts to derive potentials surfaces for New England were initiated in 1967 in conjunction with a studio in the City Planning Department of the Harvard Graduate School of Design. This work was expanded and refined as part of this demonstration grant. The emphasis in this latter work has been on correlating air pollution and potentials data, and on investigating the use of potentials in the definition of air quality control regions.

Surfaces of population density, income density, population potential, and income potential were all examined. Although, as discussed in the previous section, it is thought that the best correlations will be found with income potential, the other surfaces have significant uses:

i) population density is a measure of population-related activities such as space-heating and transportation;

ii) income density correlates remarkably well with power and energy demands on a state basis; and,

iii) where space-heating and automobile pollution are of the greatest concern, population potential may be a better measure than income potential. This depends on how much dwelling unit size and automobile usages are related to income for a particular area.

2.2.1. Calculating Potentials

The available data for calculating potentials for New England included population and effective buying income figures for the 67 counties and the 167 largest cities.[9] In addition, since

potentials reflecting the entire country (with the local variations for New England superimposed in more detail) were desired, eleven more data points were used, representing the metropolitan New York, Philadelphia, Albany and Montreal areas, each of several eastern states, and the remainder of the country divided into several regions. This allocation of the "rest of the world points" appeared to adequately show the general trend of the New England area on the national income potentials map and the detailed variations within southern New England itself.

Since sufficient air pollution data are available for only southern New England, no other areas were mapped, although data for all of New England was used in the calculations of potentials. For southern New England there are data for 27 counties and 132 cities, distributed as follows: Connecticut - 8 counties and 40 cities; Massachusetts - 14 counties and 79 cities; and, Rhode Island - 5 counties and 13 cities. It should be pointed out that there is an inconsistency in using both county and city data in the same system, especially when all of the data are to be related and a continuous surface derived from the results. Until enough data on a city basis can be obtained, there is little choice but to use this technique.

An adjustment was made in the data base as follows: the city data were subtracted from the county data for those cities contained in a particular county--if separate data were to be used for the cities. Thus, the counties actually used appear to be relatively 'rural' in nature and thus significantly inconsistent with the city data. (See Figure G-4) This technique was necessary to avoid the double counting of the city data.

Theoretically, the determination of potentials should use increments of population or income occupying very small unit areas. The potential would thus be the integration of the contribution from all the unit areas to a point and would generate a smoothed continuous surface when the potentials are calculated at all points. In practice, finite areas for which data are available must be used, rather than unit areas.

Generally, the potentials are calculated at the precise points where the data are known, but since the potentials conform to a continuous surface and incorporate the influence at a point of all other points, the potentials can be calculated at any point and not just the data points. One such alternative method is to calculate potentials on a grid basis. When potentials are calculated on a grid, a smoother surface is generated, since the grid locations do not necessarily correspond with the highest and lowest data points. These points are where the highest and lowest potentials would be expected to occur. (See Figure G-6)

Depending on how 'coarse' a grid is used, this distortion from the smoothing may be significant. The initial grid used for southern New England had locations every 17.5 miles, with the result that significant distortions occurred in the vicinity of Boston, Providence, and Fairfield County, Connecticut (suburban New York). These are areas where the gradient of potential is quite steep and subject to the local variation of numerous data points. To counteract this distortion, the potentials were recalculated and mapped with a finer grid, approximately 8.75 miles between centers. (See Figures G-7 and G-8) On the regional scale shown here, this should not produce significant distortion, although there will still be some smoothing in the vicinity of the natural peaks and pits (the highest and lowest areas, respectively).

Determining a surface of potentials involves calculating the ratio (P/r), with 'r' being the distance from a data point to the point where the potential is being determined. When this distance is close to zero (which results when the 'self-potential' is being calculated for the data point itself; in the case of potentials calculated on a grid, this can occur when the locations are very close to each other), a different estimating technique must be used. An effective distance is determined between the data point and the point where the potential is being calculated.[10]

2.2.2. Correlation Studies

Due to the lack of data very few statistical correlations were performed on the southern New England data. With the aid of standard computer programs[11] regressions were run between income density and income potential for New England. This particular set of variables and their relationship had been well documented for different periods of time, and on a national level. The main purpose in carrying out these correlations for the New England area was to examine the potentials relative to the national data (especially in terms of the inconsistency of using both county and city data). We also wanted to examine the various sub-groupings that the data might fall into: counties, cities, rural, urban, northern New England, southern New England, and ultimately high pollution and low pollution areas.

Only a few of these sub-groupings were actually done: counties separately and the entire set of data points as a whole. It does appear that logical sub-groupings do exist and that this research warrants further effort and examination. In addition,

in the same way income density has been correlated with income potential, so can air pollution data be correlated with the potential surfaces, using the same techniques.

2.2.3. Critical Features of Surfaces

A program was developed[12] at the Laboratory for Computer Graphics and Spatial Analysis which will determine the critical points and lines of a surface. These critical points include the peaks (high points), pits, (low points), ridge lines (which connect the peaks), course lines (connecting the pits, and analagous to the 'rivers' of a river basin system), and the passes and pales where the ridge lines and course lines intersect each other. Utilizing a computer map of the potentials surface, the computer attempts to determine these critical features, accomplishing this to a fairly high degree of accuracy. At the edges of the map, (see Figure G-10) there is some distortion, but in the general area of interest--from Boston to New York, and in the Providence, Worcester, Hartford, and Springfield areas, there is good detail.

There are interesting differences between the case where a grid is used to calculate potentials and where the actual data points are employed for this purpose. In the former case, some distortion in the central portion of the region as well as at the boundaries results. This is due to the coarseness of the grid, it is presumed. However, a much clearer map of critical points is produced since the surface utilizing the grid is smoother and therefore fewer local critical features appear. Thus, less is left to the judgement of the user as to which are the critical features; decisions by the user demand a fairly good knowledge of the region and the behavior of its data. It would seem from

the comparison of these two approaches, that a finer grid should combine the best aspects of the two differing results. The 8.75 mile grid was the compromise reached.

This 'watersheds' program can be used with any type of surface that would naturally exhibit critical features (as opposed to a relatively flat plane, where critical features would not be significant, and the program would distort any results). At the present time, it has been used only in conjunction with potentials surfaces since they lend themselves well to this type of analysis, but the same techniques can readily be applied to the 'surfaces' of air quality data.

A technical discussion of this program is found in Appendix 2 of this report.

2.3. AIR QUALITY CONTROL REGIONS

It was hypothesized that at a certain level of potential--associated with some determinable level of air quality--the logical boundary of an air quality control region would exist. Such an index as potential--correlating with air pollution data and being a socio-economic variable--appears to satisfy both criteria adjudged to be important in delineating regions. These criteria are air quality considerations and socio-economic factors and trends. The latter (socio-economic trends, such as population and income growth) can easily be determined with potentials. Growth areas are inherently depicted, especially when the surface is correlated with income density or population density. Secondly, if projected populations or income are used to determine the potentials, a surface of future conditions is generated. The final consideration included in delineating regions--jurisdictional boundaries--would be used to exactly fix the outline of the air quality control region, in the neighborhood of the appropriate potentials contour.

The hypothesis was tested for the Boston region at the time when the boundaries of the air quality control region were being determined. Although insufficient data were available to draw definitive conclusions, the findings were encouraging, in terms of the usefulness of the techniques. Initially potentials were calculated at the data points and on a grid with a spacing of 17.5 miles. The results were compared with the boundaries of the existing Metropolitan Air Pollution Control District (MAPCD) and the larger proposed Boston Air Quality Control Region.

When the map based on actual data points was examined (Figures G-4 and G-6) it was found that the contour line represent-

ing an income potential of 2000 x 10^6 $/mi. (herein, referred to as the "2000" contour) fairly well approximates the present MAPCD boundaries. Additional areas to the west are included in the 2000 contour, but not in the region. Conversely, parts of the region to the northeast and south are not included within this contour line.

The 1700 contour most closely approximates the then proposed air quality control region. Some significant fluctuations appear here, however, most notably in the northeast and southeast portions of the proposed region. There are several explanations why these areas are included in the region, but do not fall within the 1700 contour: (1) the potentials do not correlate well with air quality data, or regions; (2) the contours of potential due to the boundary effects of the ocean tend to close, whereas actual air pollution data extends up and down the coast to the northeast and southeast; and (3) the region should more closely approximate the 1700 contour of potentials in these areas, rather than the planning area jurisdictions used as criteria. To the west, an area of the region is not included in the 1700 potential. This, again, is probably a jurisdictional consideration based on growth trends for specific towns in this area. The 1700 contour does bulge westward at this point, as does the air quality region, but the magnitudes of the deflection here are different. More careful specification of the contours here might improve the correlation; use of a fine grid for calculating potentials could solve this problem.

Finally, to the northwest of Boston there is a significant area included in the 1700 potential which is not a part of the proposed air quality region--specifically a major part of the Lowell and Lawrence S.M.S.A.'s. The Brockton S.M.S.A. was

included in the Boston region, whereas Lawrence and Lowell were not, due to considerations of looser socio-economic ties to the Boston area and the fact that a separate air quality control region was to be designated here in the future. The decision to include Brockton and not Lawrence and Lowell is, quite naturally, based on some subjective cut-off criteria. A closer examination of the potentials surface here may well lead one to propose that Lawrence and Lowell should have been included as logically as Brockton. On the other hand, comparison of the 1700 contour and the 1600 contour show that the Worcester and Providence areas are much more reasonably considered as separate areas.

Different equi-potential lines connote different levels of wealth, activity, or influence, and quite probably different levels of the type of activity that would generate air pollution. The 1600 contour depicts the major areas of this part of New England that would be affected by adverse air quality, on a coarse level of examination. It also includes and may logically coincide with the combined air quality control areas of Boston, Providence, Worcester, and Lawrence-Lowell. This is yet another level of analysis of the correlation of potentials and air quality control regions.

When we pick out the critical features on the surface we find that three peaks are exhibited: Boston, Providence, and Worcester, in descending order of importance. The contour lines symbolically separate the hills surmounted by these three peaks. When these 'hills' are related to air pollution data as hypothesized here they logically divide the region into three areas. The course line to the south and west of Boston follows quite closely the boundary of the proposed air quality control region. The major difference is that the proposed area is smaller than the course line would suggest.

For the region to more nearly coincide with the course lines, all towns in this area would have to be contained in either the Boston, Providence, or Worcester air quality region. In other words, the use of the course lines to define boundaries cannot implicitly take into account 'clean' areas that would be between regions; all areas must be in one region or another. However, using specific contours as previously explained can handle this type of situation.

There are several other points to be made from this mapping of critical features. First of all there is no course line to the north of Boston to indicate either a division between the Boston and Lowell-Lawrence areas, or an end to the Boston region. In the first instance, it strengthens the case for including Lawrence and Lowell in the proposed Boston air quality control region. In the second case it points up the fact that the course lines only divide regions; they do not define the outer boundary of some logical region except in terms of the relationship to the adjacent region. Here the program finds no adjacent region to the north.

The potentials were also examined for a grid with a 17.5 mile spacing but the results were less conclusive. The contour lines do not conform to the boundaries of either the present or larger proposed region nearly so well as in the case where the actual data points were used to calculate the potentials.

The peaks have definitely been smoothed out with the use of a grid. Whereas the 2000 contour fell predominantly outside the MAPCD boundary before, now it is mainly inside this boundary. On the other hand the 1700 contour is in much the same location as before, although far more concentric (again a property of the smoothing of the surface), and does not

accentuate the Lowell-Lawrence area as much. The 1600 contour in the grid case is just outside the proposed region, whereas previously it included the Providence and Worcester areas as well. Clearly at this scale, the grid system does not appear to give any meaningful results, either positively or negatively. In addition the critical features are not clearly revealed.

Continued experiments with more detailed grid spacing in combination with the actual data points would quite probably give useful results. The use of critical features in particular may prove to be a powerful tool for regional definition of control and standard-setting areas.

REFERENCES

(1) These include:
Warntz, William, Macrogeography and Income Fronts, Regional Science Research Institute, Monograph Series Number Three, Philadelphia, 1965

Stewart, J.Q., and W. Warntz, "Some Parameters of the Geographical Distribution of Population", Geographical Review, Vol. 49, 1959

(2) 1967 income potentials were derived from county effective buying income data from Sales Management, Inc, New York, in computer useable form. Locations reflected the town of the county court house, rather than the centroid of the county.

(3) David H. Slade, "Modeling Air Pollution in the Washington D.C. to Boston Megalopolis", Science, Vol. 157, No. 3794, September 15, 1967, pp. 1304-1307.

(4) Public Law 90-148; 90th Congress, S. 780, November 21, 1967; Air Quality Act of 1967;
Federal Register, Vol. 33, No. 10; Tuesday, January 16, 1968. "Definition of Atmospheric Areas" pp. 548-549

(5) Under the direction of William Warntz at the Laboratory for Computer Graphics and Spatial Analysis, Robert Mercready prepared a detailed map of the United States which can display potentials data for the more than 3000 counties of the country, based on 1967 population and income data. This base map was used for the data shown in Figures G-1 and G-2.

(6) 1967 National Air Sampling Network Data, U.S.P.H.S., Cincinnati

(7) These studies were undertaken from 1967 thru 1969 as part of course work in the Harvard Graduate School of Design.

(8) The data were gathered by 24-hour hi-vol samplers.

- 1966 NASN and CAMP data, U.S.P.H.S., Cincinnati.
- The Commonwealth of Massachusetts "Special Report on the Investigation and Study of Air Quality in the Metropolitan (Boston) Air Pollution Control District, 1965-1966; September 16, 1968, Air Use Management Program, Bureau of Environmental Sanitation, Department of Public Health; also,
- Updates and estimates done in conjunction with Bureau personnel
- Connecticut, Department of Health
  Connecticut Air Sampling Network Data 1966-1968

(9) These data were from
Sales Management, Inc., New York. 1967 City and County Data For New England on Population, Effective Buying Income, Per Capita Income, Retail Sales. 1967 County Data on Employment, By 2-Digit S.I.C. (in computer useable form).

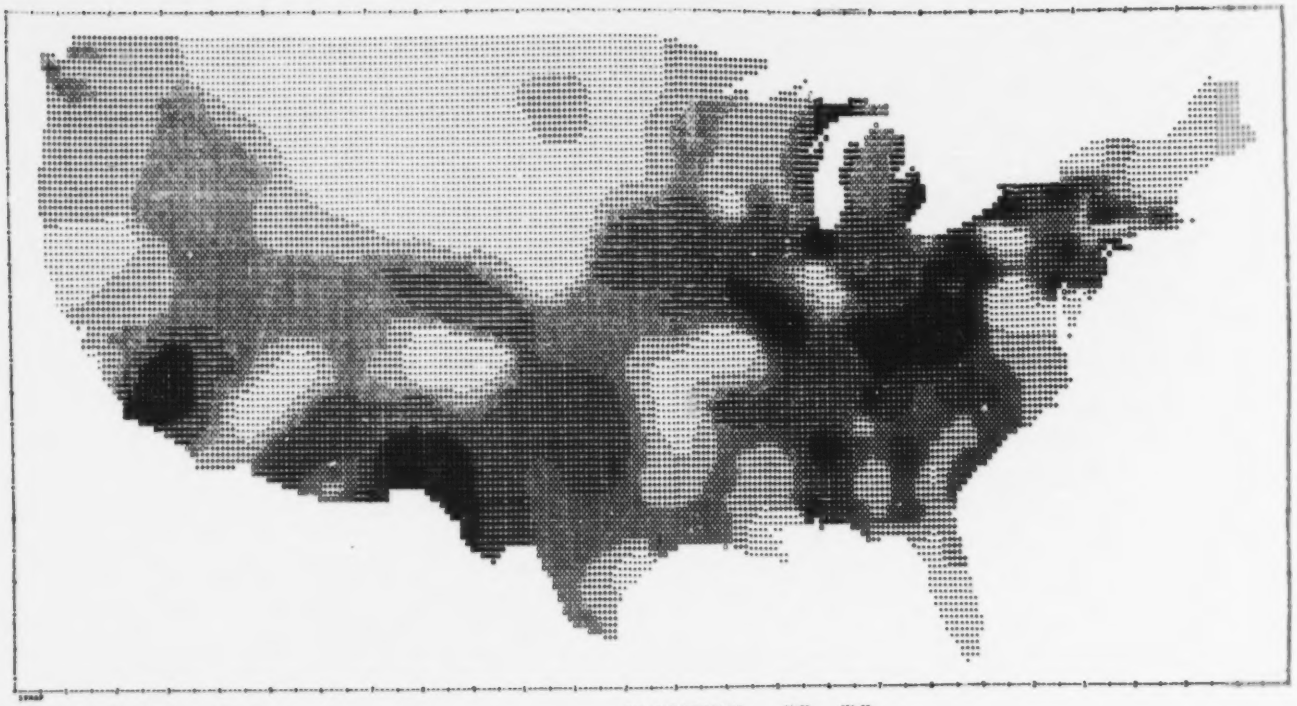
(10) The procedure used here was developed by William Warntz: Use the area associated with the data point (the area of the city or county, herein); assume this area is in the shape of a circle, and use one-half the radius of the circle as the appropriate distance for the calculation of potential. Although this does not affect the general potential surface for a region or area, it does affect the relative values of the local peaks; this should be borne in mind when analyzing the potentials surfaces, especially when the grid procedure is used. At this time, there is no way to verify the validity of this estimation method.

(11) The University of California Bio-medical programs (BMD) were adapted for this study.

(12) The "watersheds" program was developed by Donald Shepard in 1967 and refined from 1968 thru 1969.

APPENDIX ONE

COMPUTER GRAPHICS



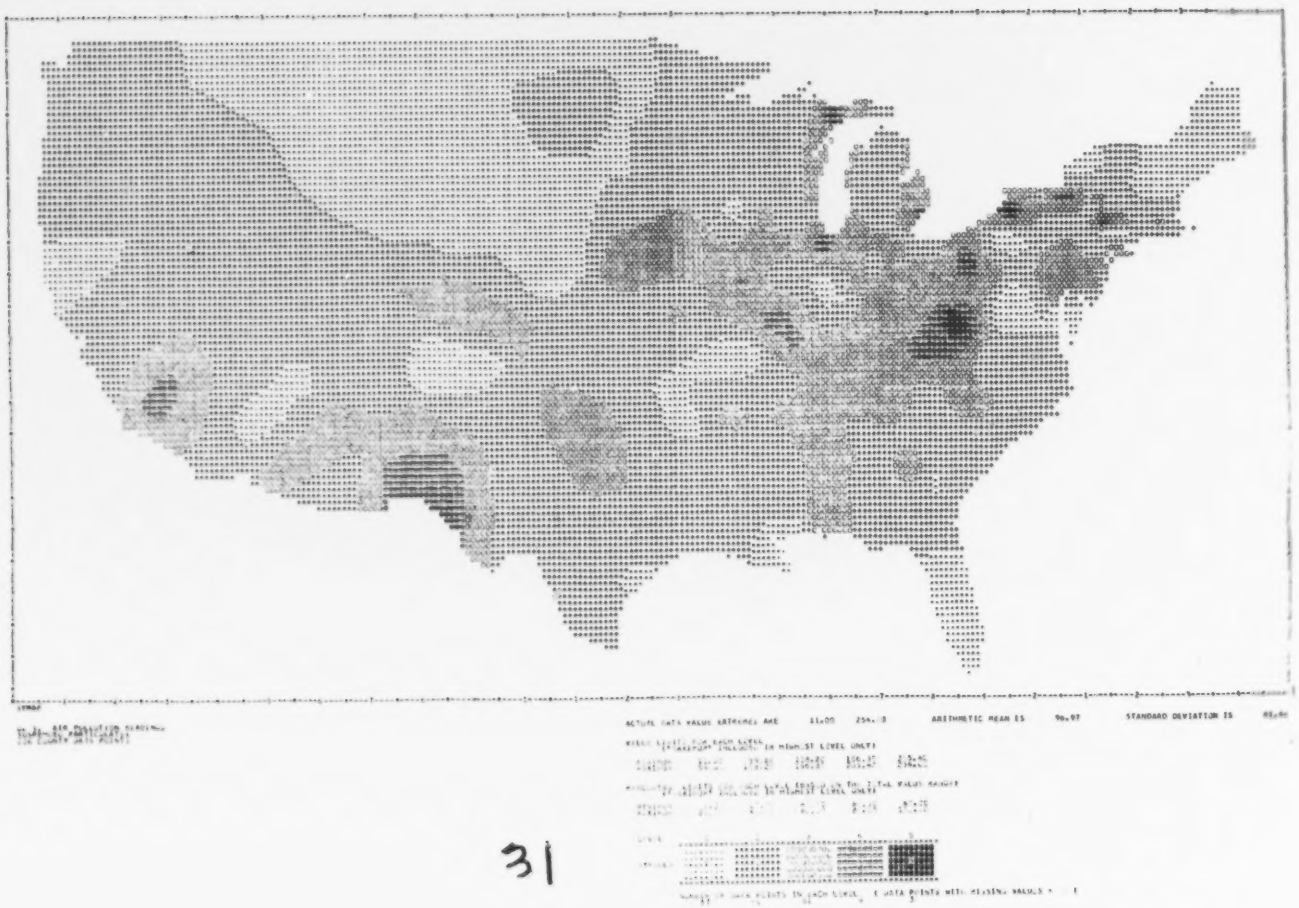
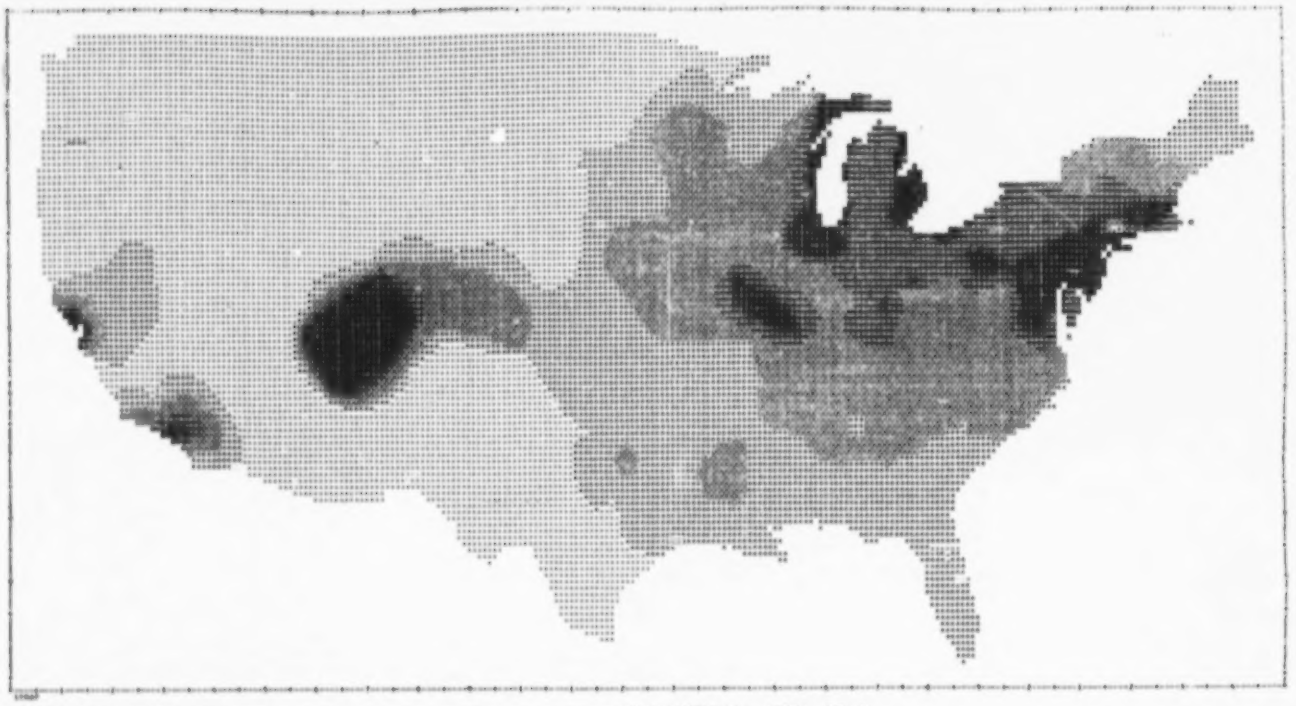
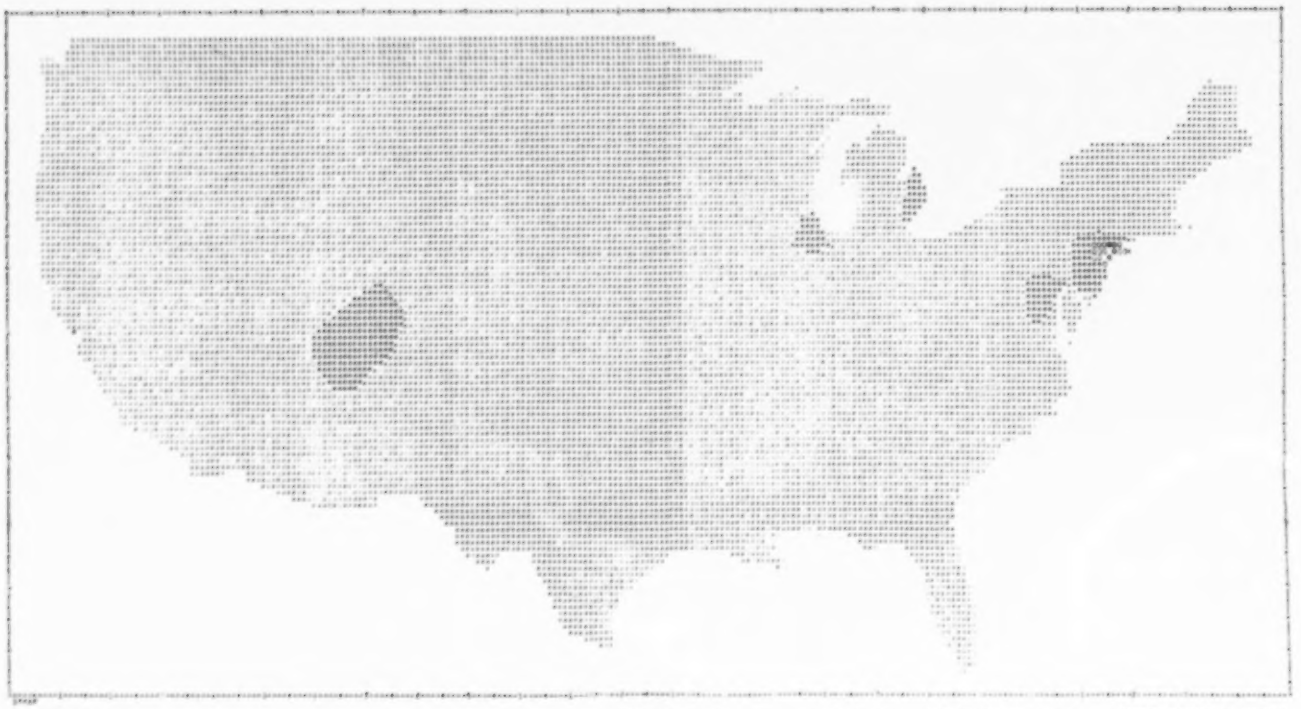


FIGURE G-1

SOUTHERN NEW ENGLAND STUDY
Concentration of Suspended Particulates
Based on 206 Selected Urban and Rural Recordings
a. Equal Number of Data Points in Each Level
b. Equal Interval Contour Levels



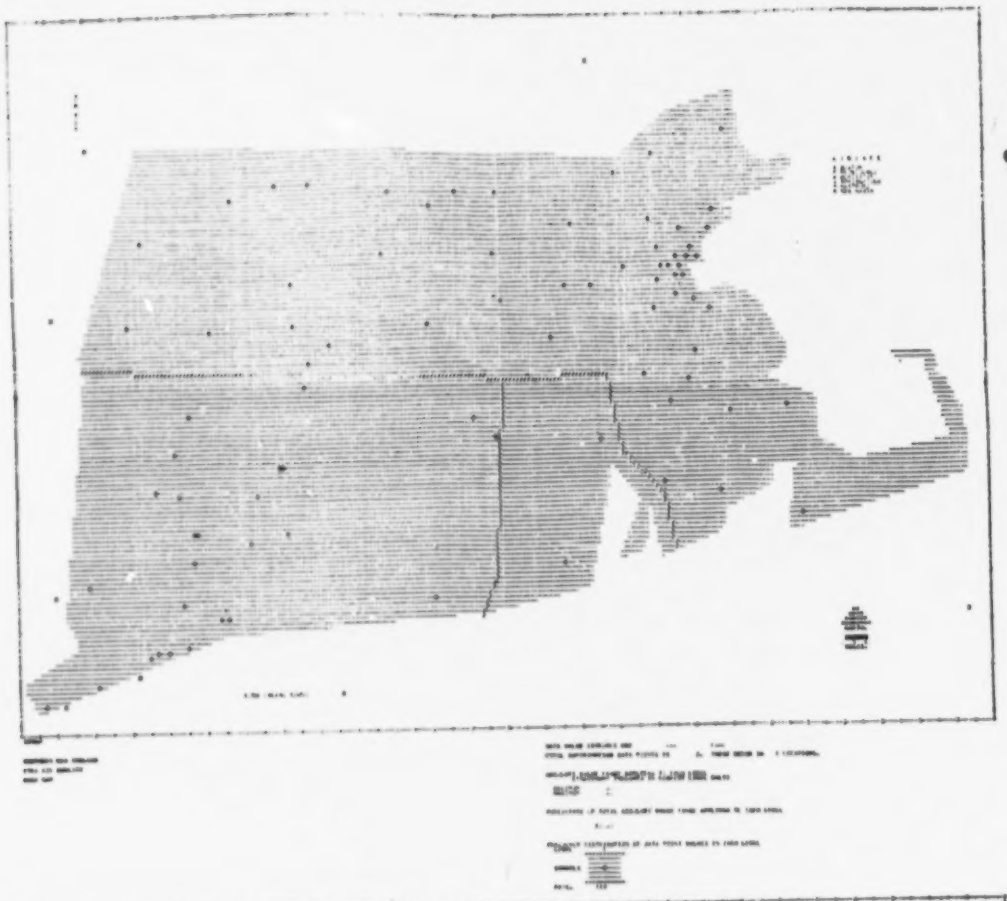


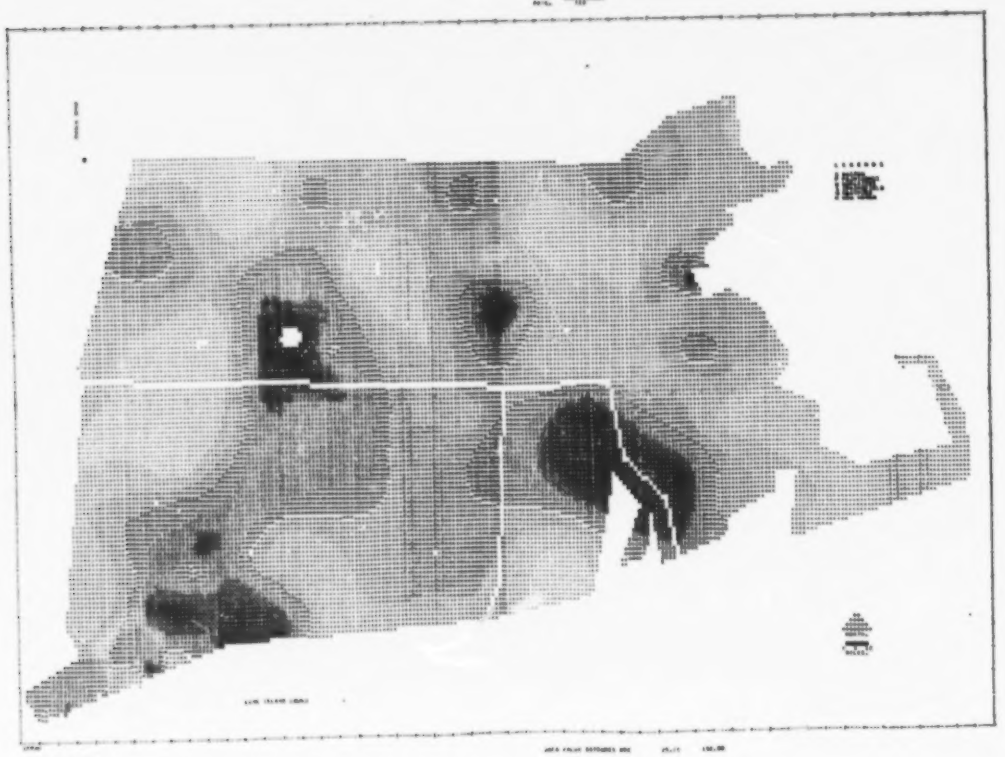
Details ... in this document ... be better studied on microfiche

FIGURE G-2

SOUTHERN NEW ENGLAND STUDY
United States Income Potential
Based on 206 Urban and Rural County Values
a. Equal Number of Data Points in Each Level
b. Equal Interval Contour Levels

Details of illustrations in this document may be better studied on microfiche





93

FIGURE G-3

SOUTHERN NEW ENGLAND STUDY
Concentration of Suspended Particulates
a. Base Map of Monitoring Stations
and Background Data Points
b. Based on Actual and Estimated Values

Details of illustrations in this document may be better studied on microfiche

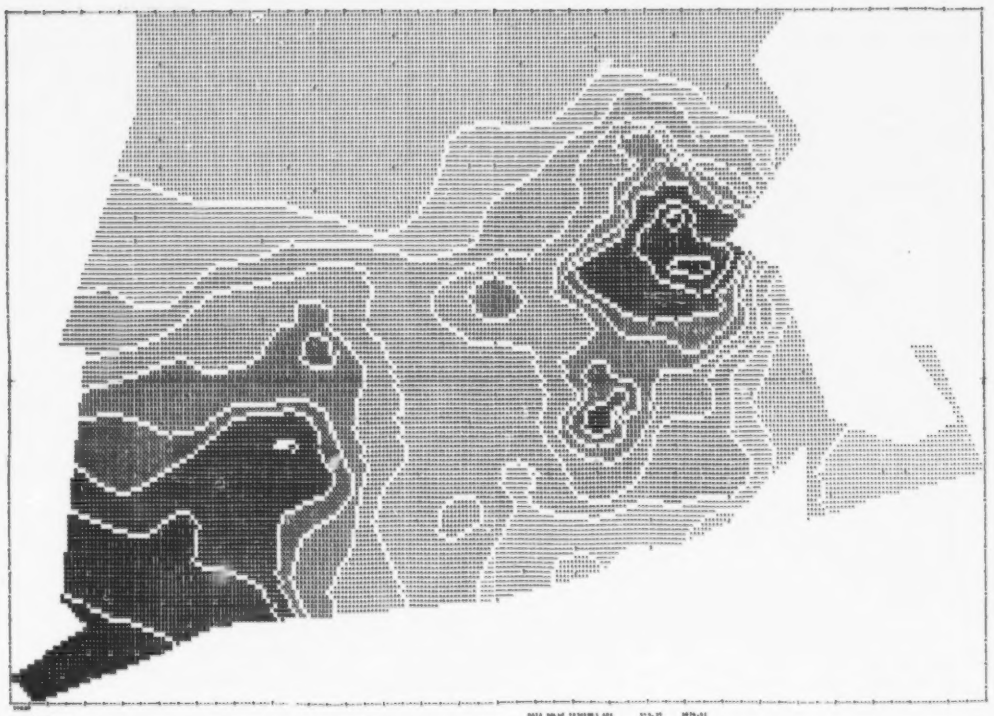


FIGURE G-4

SOUTHERN NEW ENGLAND STUDY
Income Potential Calculated at City & County Data Points
a. Surface Based on Counties Only
b. Surface Based on Cities and Counties

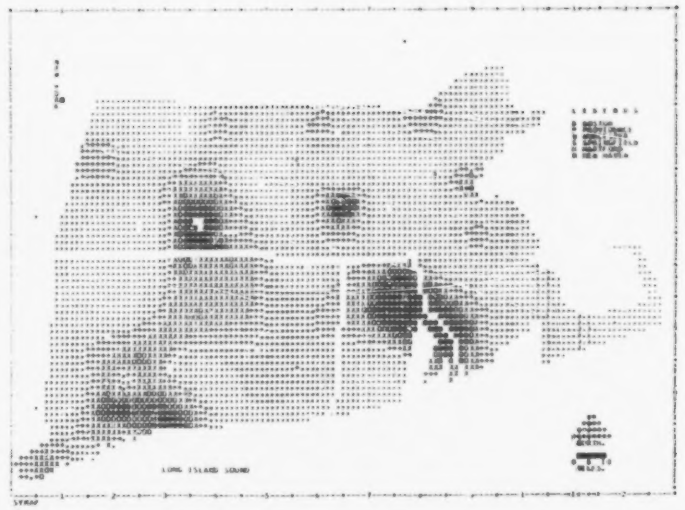
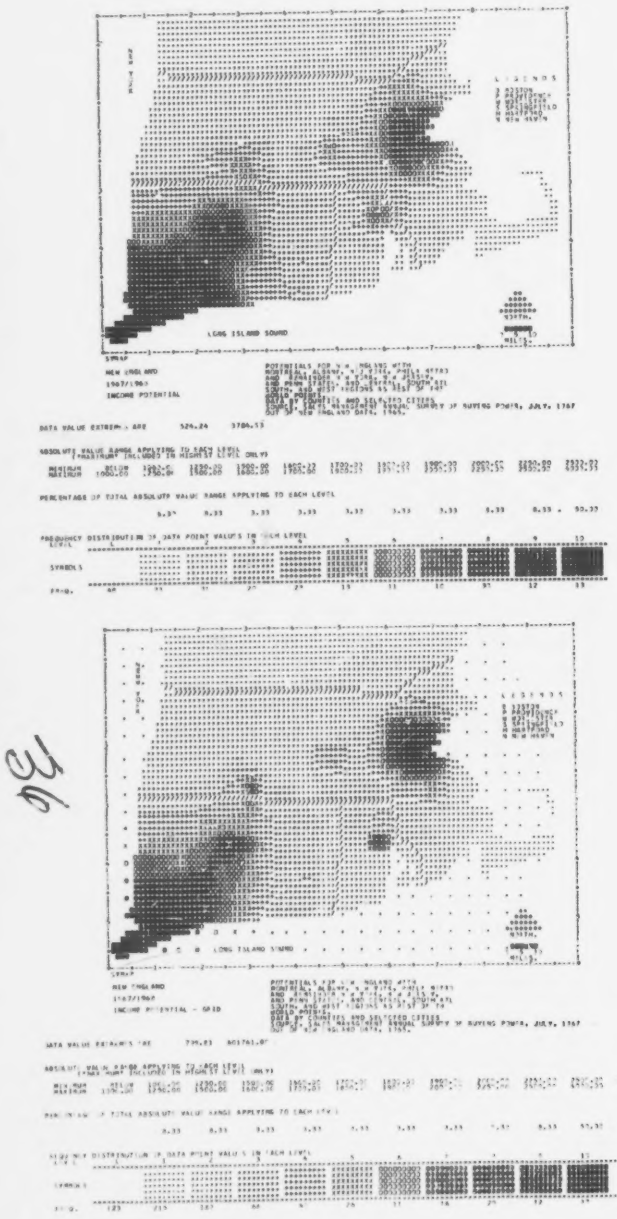
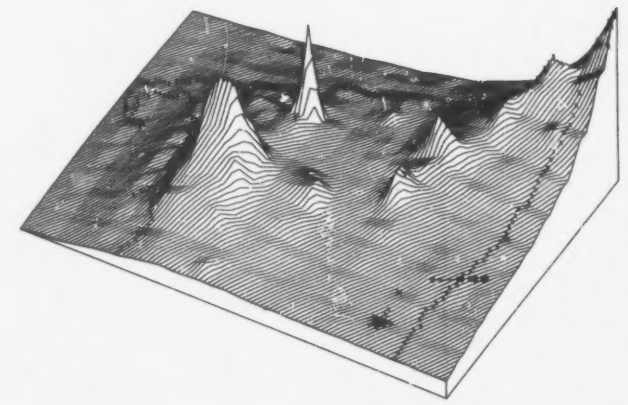


FIGURE G-5

SOUTHERN NEW ENGLAND STUDY
Concentration of Suspended Particulates
a. Based on Measured 1966 Values Only
b. Based on Actual and Estimated Values

Depts. of Health, Education & Welfare





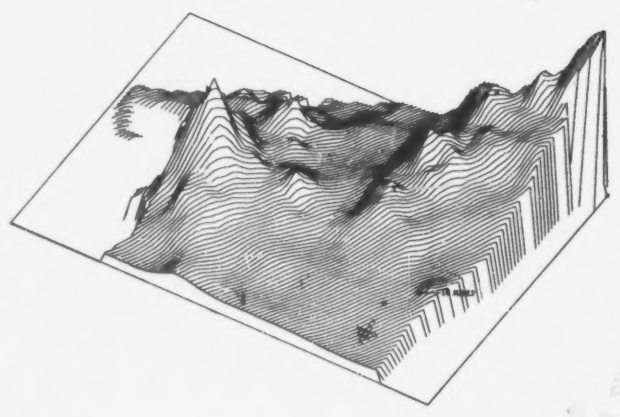


FIGURE G-6

SOUTHERN NEW ENGLAND STUDY
Income Potential Surface
a. Calculated at City and County Data Points
b. Oblique View of a.
c. Calculated on an 8-mile Grid
d. Oblique View of b.

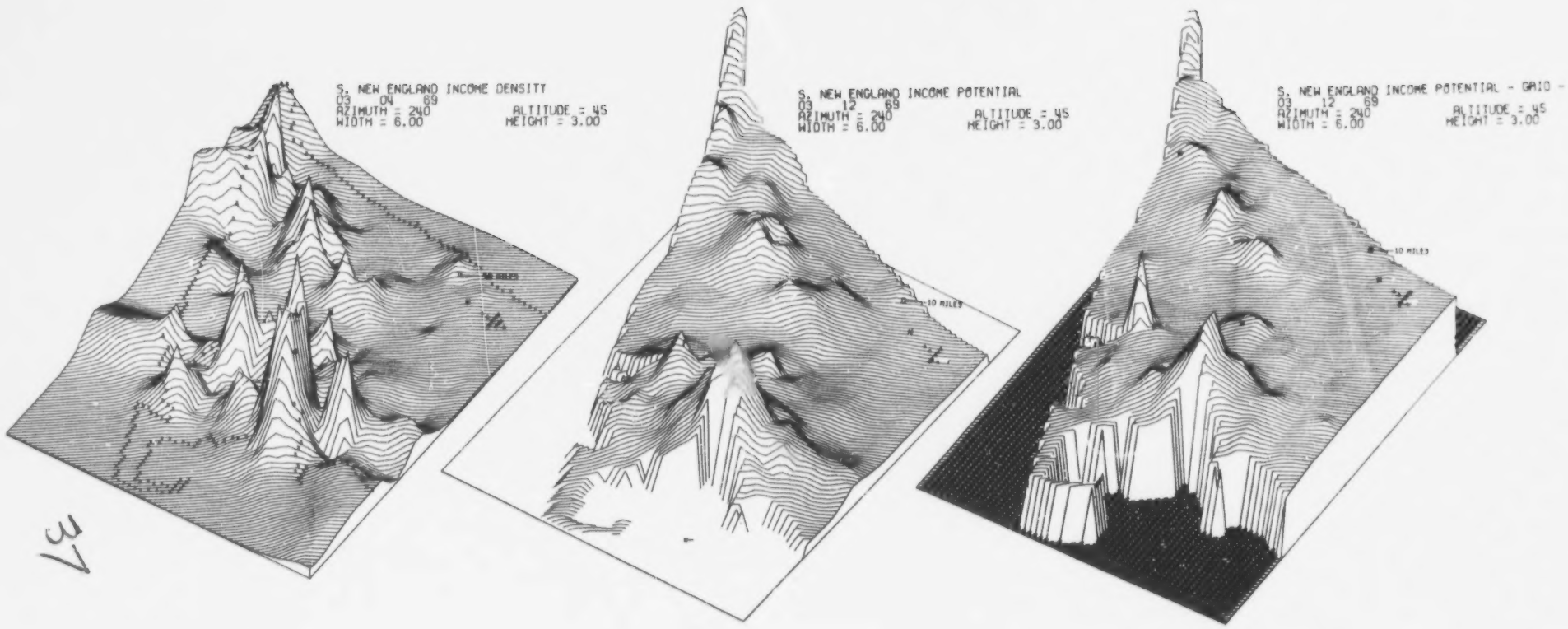


FIGURE G-7

SOUTHERN NEW ENGLAND STUDY
Comparison of Income Surfaces — Oblique Views
a. Income Density
b. Income Potential at Data Points
c. Income Potential on an 8-mile Grid

Details of illustrations in this document may be better studied on microfiche

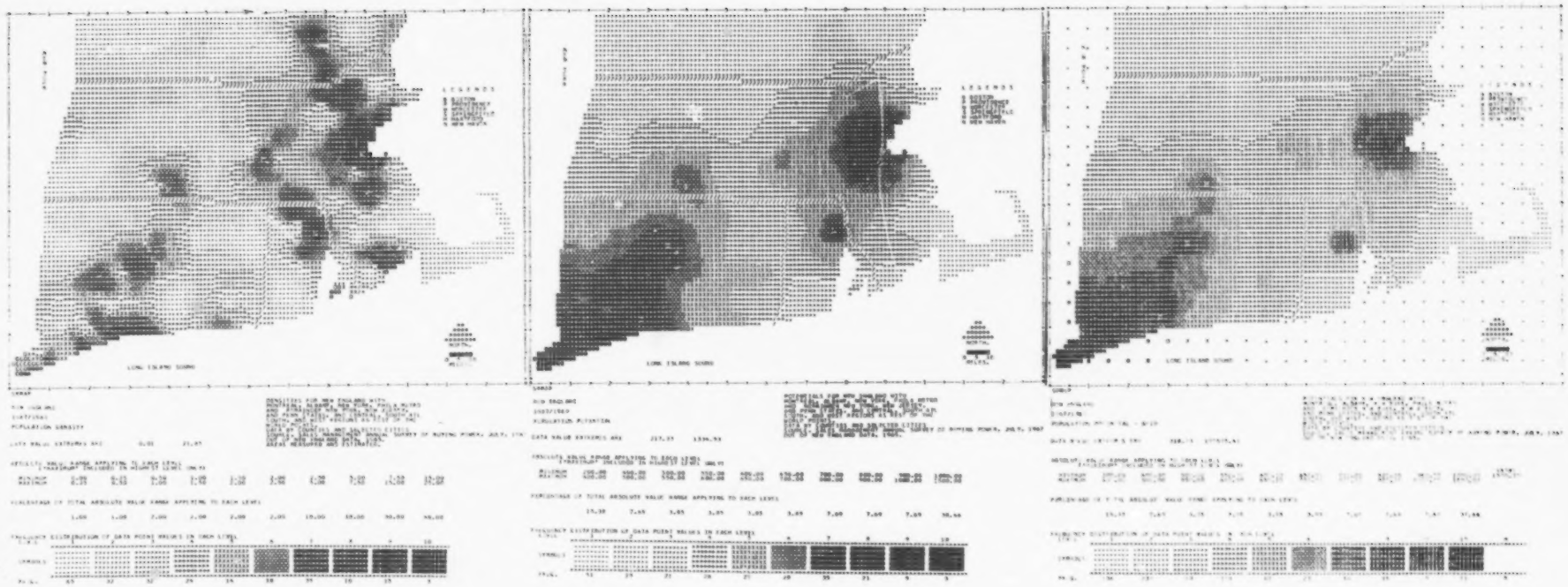
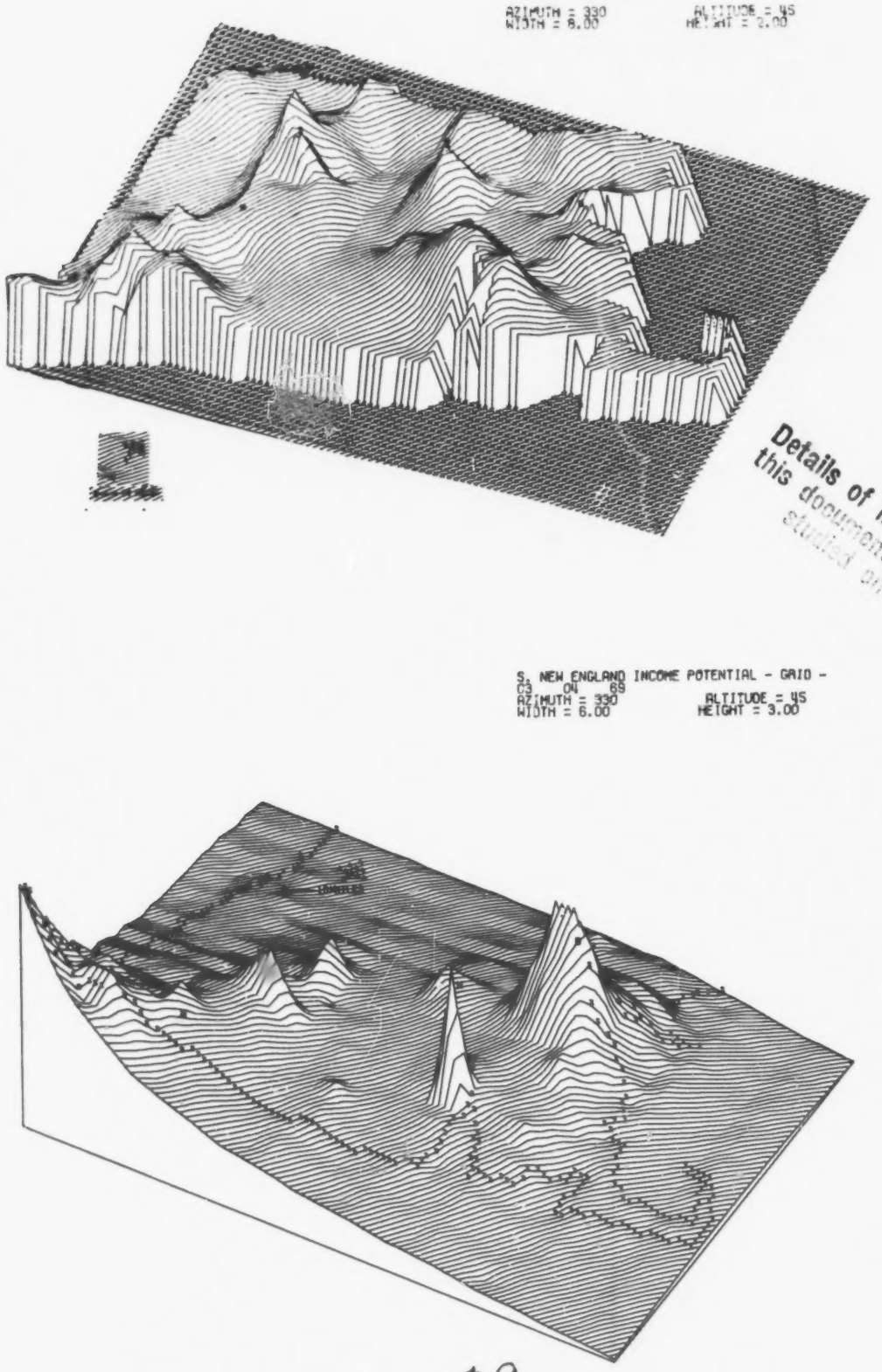


FIGURE G-8

SOUTHERN NEW ENGLAND STUDY
Comparison of Population Surfaces
a. Population Density
b. Population Potential at Data Points
c. Population Potential on an 8-mile Grid

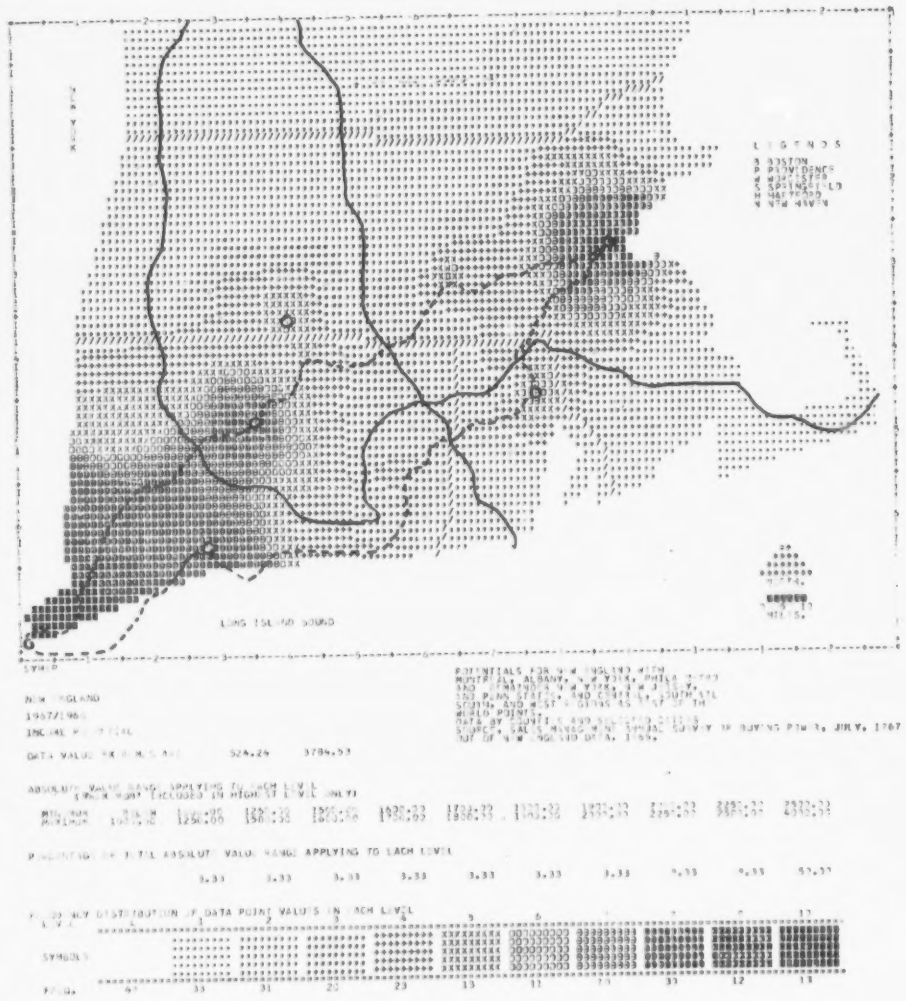
Details of illustrations in this document may be better studied on microfiche

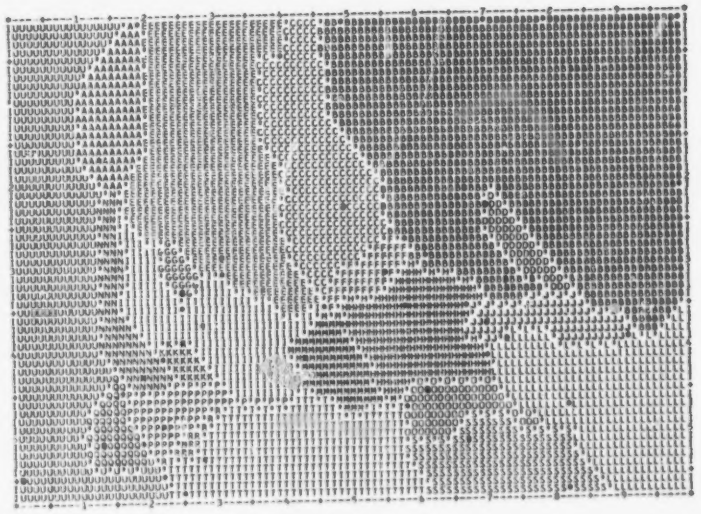


Details of illustrations in this document may be better studied on microfiche

FIGURE G-9

SOUTHERN NEW ENGLAND STUDY
Comparison of Air Pollution and Income Potential Surfaces
Oblique Views
a. Concentration of Suspended Particulates
Actual and Estimated Values Used
b. Income Potential — Calculated on an 8-mile Grid





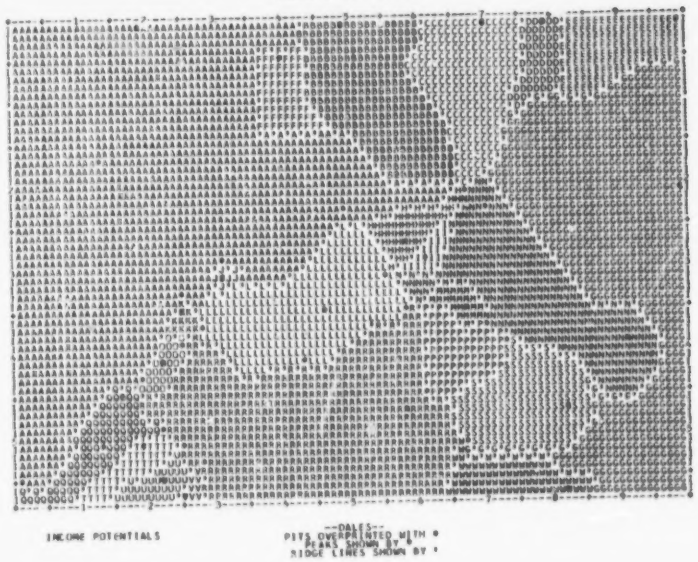


FIGURE G-10

SOUTHERN NEW ENGLAND STUDY
Critical Features of Income Potential Surface
a. Income Potential Surface
b. Hills and Course Lines
c. Dales and Ridge Lines

Details of illustrations in this document may be better studied on microfiche

APPENDIX 2

CRITICAL FEATURES PROGRAM

The "watersheds" program was written at the Laboratory for Computer Graphics and Spatial Analysis in 1967 and 1968. Since no recent major changes have been made in the program the following write-up refers to an IBM 7094 version of the program. Conversion to the IBM 360 or other computers is possible with minimal re-programming.

A. Main Features of the Program

The program uses as input a surface created by SYMAP or some similar computer mapping program. Given the quantitative surface, the program determines and displays as maps:

(i) magnitude of gradient,
(ii) direction of steepest descent or ascent, and,
(iii) hills and dales with course and ridge lines.

Magnitude of gradient is estimated by the average rate of change of map values from adjacent character locations above and below, and to the sides, of each character space. The scale divisions for levels two through nine proceed in eight equal increments to the maximum gradient designated. Steeper gradients are designated by level ten. Critical points, at which the gradient is less than 2% of the maximum in magnitude, are assigned to the first level.

Direction of steepest descent, or ascent, is determined as follows. At each character, the computer examines the average slope to each of the adjacent eight characters, skipping

A1

over spaces containing background or border. If there is no direction of steepest descent (ascent), the space is coded as a peak or pit. If an area is level, one or more points in its lower right is arbitrarily designated as a pit (peak).

To determine hills and dales, the computer begins at each pit (peak) and spirals outward. At each character location, it examines the direction of steepest descent to see whether that location leads directly to the pit (peak) under consideration, or leads to a location already assigned to the respective dale (hill) and thus indirectly to the pit (peak) under consideration, or leads to some other dale (hill). Pits (peaks) are overprinted with an asterisk. Dots are placed at the boundary of hills or dales to show ridge or course lines. Finally a list of pits or peaks is printed. The analysis is performed in 48 row sections. Missing letters among pit (peak) symbols are those used in linking the sections.

B. Program Requirements

The input data consists of:

1. a binary magnetic tape with map values, as prepared by SYMAP or another compatible program,

2. certain control cards, governing which one or more of the following analyses are produced:

*map showing magnitude of gradient, the steepest gradient appearing darkest.
*chart explaining symbolism.
*map showing direction of steepest descent.
*map showing dales (watersheds) with pits and separating ridge lines.

#2

*list of pits containing symbol, location, and area of corresponding dale.
*map showing direction of steepest ascent.
*map showing hills (tributary areas to peaks) with peaks and separating course lines.
*list of peaks containing symbol, location, and area of corresponding hill.

The program is written in FORTRAN IV, and has been run on the IBM 7094. The program consists of twenty-three decks (a main routine and twenty-two subroutines) and numbers about nine hundred cards. The program requires about 15K words for the object decks, plus an additional 9K for utility and systems routines. The program fits easily into the 7094 without overlays and with ample space for large buffers.

Additional requirements are the availability of an intermediate scratch tape (logical tape 4) and a binary input tape (logical tape 8). Random access disk is also needed, the amount depending on the size of the map being analyzed. A map of about five square feet, or 480 rows, would require about 500 disk tracks, each of at least 130 words. In order to access disks from the FORTRAN program, several utility routines obtained from the SHARE library, and part of the System library at Harvard, are used. They are subroutines DPUT, DTAKE, and DCHEK in deck F4DK.

The program is written to be compatible with SYMAP Version 5. Output from the SYMAP program is saved for analysis by this program through SYMAP elective 21. A tape produced by some other program, in the same format, may also be used as input.

43

The first record of the input tape contains the following words: NROW, NCOL, RINCH, CINCH, ZMIN, ZMAX, where

NROW is row number of bottom border,
NCOL is column number of right border,
RINCH is number of rows per inch (normally 8)
CINCH is number of columns per inch (normally 10),
ZMIN is lowest data value, and
ZMAX is highest data value.

For each map strip there are NROW-1 records of identical format, one record for each row within the borders. Each word in these records is the floating point value at the corresponding character. This information on the binary tape is also in the form required as input by the SYMVU program.